TWO-DIMENSIONAL-ORIENTED LINEAR DISCRIMINANT ANALYSIS FOR FACE RECOGNITION

Muriel Visani[1], Christophe Garcia[1] and Jean-Michel Jolion [2]
[1]*France Telecom R&D DIH/HDM*
4, rue du Clos Courtel
35512 Cesson-Sevigne, France
{muriel.visani,christophe.garcia}@rd.francetelecom.com
[2]*Laboratoire LIRIS, INSA Lyon*
20, Avenue Albert Einstein
Villeurbanne, 69621 cedex, France
Jean-Michel.Jolion@insa-lyon.fr

Abstract In this paper, a new statistical projection-based method called Two-Dimensional-Oriented Linear Discriminant Analysis (2DO-LDA) is presented. While in the Fisherfaces method the 2D image matrices are first transformed into 1D vectors by merging their rows of pixels, 2DO-LDA is directly applied on matrices, as 2D-PCA. Within and between-class image covariance matrices are generalized, and 2DO-LDA aims at finding a projection space jointly maximizing the second and minimizing the first by considering a generalized Fisher criterion defined on image matrices. A series of experiments was performed on various face image databases in order to evaluate and compare the effectiveness and robustness of 2DO-LDA to 2D-PCA and the Fisherfaces method. The experimental results indicate that 2DO-LDA is more efficient than both 2D-PCA and LDA when dealing with variations in lighting conditions, facial expression and head pose.

Keywords: Two-Dimensional-Oriented Linear Discriminant Analysis, Face Recognition, Feature Extraction, Statistical projection, Two-Dimensional Principal Component Analysis.

1. Introduction

Since the seminal work of Sirovich and Kirby [6], which showed that Principal Component Analysis (PCA) could be efficiently used for representing images of human faces, statistical projection-based methods have been widely used in the context of automatic face recognition. Turk and Pentland [7] proposed the very well-known Eigenfaces method, based on PCA, where a face image can be represented as a weighted sum of a collection of images (eigenfaces) that define a facial basis. Belhumeur *et al.* [1] introduced the Fisherfaces method, based on Linear Discriminant Analysis (LDA) where class information, i.e. the identity of each face image, is taken into account for enhancing separation between different classes, while building the face space.

In PCA-based and LDA-based face recognition methods, the $h \times w$ 2D face images must be first transformed into 1D image vectors of size $h \cdot w$, which leads to high-dimensional image vector space, where statistical analysis, i.e. covariance matrix calculation and eigen system resolution, is costly, difficult and may be unstable. To overcome these drawbacks, Yang *et al.* [10] proposed recently the Two Dimensional PCA (2D-PCA) method that aims at performing PCA using directly the face image matrices, keeping the 2D structure of the face images. They have shown on various databases that 2D-PCA is more efficient than PCA for the task of face recognition. In addition, we have shown [8] on FERET [4] that 2D-PCA is more robust than PCA when dealing with face segmentation inaccuracies such as misaligned or badly scaled face images, with low-quality images and partial face occlusions.

In this paper, we propose a novel class-based projection technique, called Two-dimensional-Oriented Linear Discriminant Analysis, that achieves better recognition results than traditional LDA-based approaches by taking advantages of the 2D matrix representation of the face images while substantially reducing computational and storage costs.

The remainder of the paper is organized as follows. In section 2, we describe in details the principle and the algorithm of the proposed 2DO-LDA method, pointing out its advantages over previous projection-based methods. In section 3, a series of four experiments, on different international data sets, is presented to demonstrate the effectiveness and robustness of 2DO-LDA, with respect to variations in lighting conditions, facial expression and head pose and compare its performances with respect to the LDA and 2D-PCA methods. Finally, conclusions are drawn in section 4.

2. 2D-Oriented Linear Discriminant Analysis

The classifier is constructed from a training set of n face image matrices X_i, containing $h \times w$ pixels, labeled by their corresponding identity. The views of one person form a class. The aim is to find a projection matrix P, of size $w \times k$, providing efficient separation of the projected classes according to:

$$\hat{X}_i = X_i \cdot P \tag{1}$$

where $\hat{X}_i$ is the $h \times k$ projected matrix of X_i onto the orthonormal basis P of the projection space. Column vectors $(P_i)_{i=1\ldots k}$ of P will be referred to as *2D-Oriented Discriminant Components* (2DO-DCs) in the following.

The 2DO-DCs are chosen to jointly maximize the mean variation between classes and minimize the mean of the variations inside each class. Therefore, P can be chosen as the $w \times k$ matrix maximizing the following *generalized Fisher criterion*:

$$J(P) = \|(\hat{S_w})^{-1}\hat{S}_b\| \quad (2)$$

where $\hat{S_w}$ and $\hat{S}_b$ are respectively the *generalized within-class* and *between-class covariance matrix* of the n projected image matrices $\hat{X}_i$, defined as:

$$\hat{S_w}=\frac{1}{n}\sum_{c=1}^{C}\sum_{X_i\in\Omega_c}(\hat{X}_i-\bar{\hat{X}}_c)^T(\hat{X}_i-\bar{\hat{X}}_c) \text{ and } \hat{S}_b=\sum_{c=1}^{C}\frac{n_c}{n}(\bar{\hat{X}}_c-\bar{\hat{X}})^T(\bar{\hat{X}}_c-\bar{\hat{X}}) \quad (3)$$

where $\bar{\hat{X}}_c$ is the mean matrix of the n_c projected images of class Ω_c (among C different classes) and $\bar{\hat{X}}$ is the mean matrix of all the n projected images of the training set. According to equations (1) and (3), criterion (2) is equivalent to the following criterion:

$$J(P) = \frac{|P^T S_b P|}{|P^T S_w P|} \quad (4)$$

where S_w and S_b are respectively called the *generalized within-class* and *between-class covariance matrix* of the training set:

$$S_w=\frac{1}{n}\sum_{c=1}^{C}\sum_{X_i\in\Omega_c}(X_i-\bar{X}_c)^T(X_i-\bar{X}_c) \text{ and } S_b=\sum_{c=1}^{C}\frac{n_c}{n}(\bar{X}_c-\bar{X})^T(\bar{X}_c-\bar{X}) \quad (5)$$

where $\bar{X}_c$ and $\bar{X}$ are respectively the mean of the n_c images of class Ω_c and the mean of all the n images of the training set.

Under the assumption that S_w is non-singular, the k vectors P_i maximizing criterion (4) are the k orthonormal eigenvectors of matrix $S_w^{-1}S_b$ corresponding to the largest eigenvalues. The matrix S_w is generally invertible due to the low dimension of the 2DO-DCs relative to the number of training samples $(n \gg w)$.

Once they have been sorted in descending order from their corresponding eigenvalues, the number k of 2DO-DCs to consider can be determined as for the eigenfaces method [9], traditionally by removing a given percentage of the last eigenvectors.

As a statistical projection method, 2DO-LDA can be used for image compression, even if the projection space is chosen to be more discriminative than representative. The projected image $\hat{X}_i$ and P can be combined to obtain a reconstruction of the original image X_i ; some results are shown in Figure 1.

Classification of face images is performed in the projection space defined by P: when comparing two faces X_a and X_b, they are first projected onto P according to equation (1), giving their projections $\hat{X}_a$ and $\hat{X}_b$. Then, a matrix-to-matrix distance is calculated between $\hat{X}_a$ and $\hat{X}_b$, for instance the following

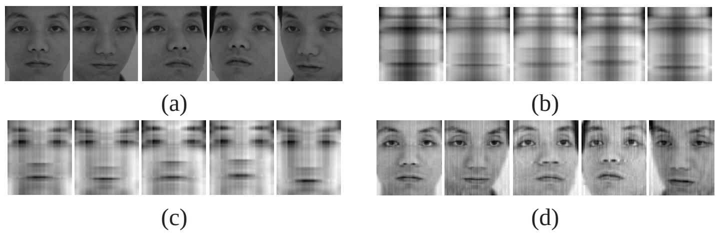

Figure 1. (a) Original images (Asian Face Database PF01). (b) Corresponding reconstructed images with $k = 2$ 2DO-DCs. (c) With $k = 3$. (d) With $k = 20$. The projection space is constructed from the training set of the first experiment (see section 3). With the third 2DO-DC the facial features (eye, nose, mouth) appear, but the head poses are not distinguishable yet. With more 2DO-DCs a good visual quality of reconstruction is obtained.

distance, used by Yang. *et al.* [10]:

$$d(\hat{X}_a, \hat{X}_b) \;=\; \sum_{j=1}^{k} \|\hat{X_a^j} - \hat{X_b^j}\|_2 \qquad (6)$$

where $\hat{X_i^j} = X_i P_j$ is the projected vector of image X_i on the the j^{th} 2DO-DC P_j, and $\|\cdot\|_2$ is the standard L_2 norm.

It can be pointed out that 2DO-LDA offers strong advantages in comparison with 2D-PCA and the usual LDA method:

- In 2D-PCA the projection space is chosen to retain most of the total scatter of the training set, no matter if that scatter is explained by variations inside the same class (variations in facial expression for instance) or between two different classes. Thus, the projection space constructed from 2D-PCA can represent noise and this method is more suited for face representation than for face classification. It will be shown in section 3 that 2DO-LDA have a stronger discriminative power than 2D-PCA;

- 2DO-LDA is numerically more stable than the usual LDA method: for the LDA method the sample images are vectors of length $w \cdot h$, and this large dimension leads to numerical instability when computing the within and between-class covariance matrices, from these vectors;

- 2DO-LDA allows an important storage gain with respect to the usual LDA method: while for the LDA method the length of the projection vectors P_i is $w \cdot h$, for 2D-PCA their length is w. Moreover, the number k of selected projection vectors for LDA is traditionally 60% of the number of samples of the training set. We will see in section 3 that the number of 2DO-DCs needed to provide good face recognition rates is much smaller;

- Concerning LDA, the length of the projection vectors $w \cdot h$ is usually much larger than the number of samples n. Therefore, the within-class covariance matrix of the training set is generally non-invertible. The trick traditionally used is to perform LDA into a subspace previously constructed from PCA [5]. The corresponding algorithm will be denoted by "PCA+LDA" in the following. Applying first PCA generates an additive computational cost and leads to a loss of information that could, if kept, be discriminative. Concerning 2DO-LDA, the S_w matrix is generally invertible and therefore the algorithm can be applied directly on the training set.

3. Experimental Results

Four experiments are performed to assess the effectiveness and robustness of 2DO-LDA with respect to variations in lighting conditions, facial expression and head pose and compare its performance with LDA and 2D-PCA.

Three face databases are used: the Asian Face Image Database PF01 [2], containing 17 views of each of 107 persons, the well-known FERET [4] face database, and the BioId Database [3], containing 1521 face images of 23 people, extracted from video sequences.

A face image preprocessing step is first applied to each image: it consists in centering the face in the image, setting the image to a size of 65 pixels wide by 75 pixels high, and equalizing its histogram.

The first two experiments provide a comparison of the robustness to variations in head pose and facial expression, the third experiment aims at evaluating the efficiency in the presence of illumination variations. The last experiment consists in matching video sequences of faces with different lighting conditions, head poses and facial expressions. Figure 2 shows samples of the training and test sets used for these experiments.

Experimental results are analyzed through two graphics: the compared recognition rates of 2D-PCA and 2DO-LDA across a varying number of projection vectors k (first column of Figure 3) and the compared Cumulative Match Characteristic (CMC) curves for LDA, 2D-PCA and 2DO-LDA (second column of Figure 3). A face is said to be recognized at rank j if an image of the same person is among the j^{th} nearest into the projection space. The distance (6) is used for 2D-PCA and 2DO-LDA. Concerning LDA a L_2 distance is performed in the projection space. In each CMC curve, the number k of projection vectors has been chosen so as to maximize the performances of the algorithm.

The first experiment (see Figure 2.a) is performed on the Asian Face Image Database PF01. The training set contains 535 images of faces, 5 views in near-frontal pose per person. The test set contains 428 images (4 views per person), with stronger non-frontal head poses than in the training set. The training

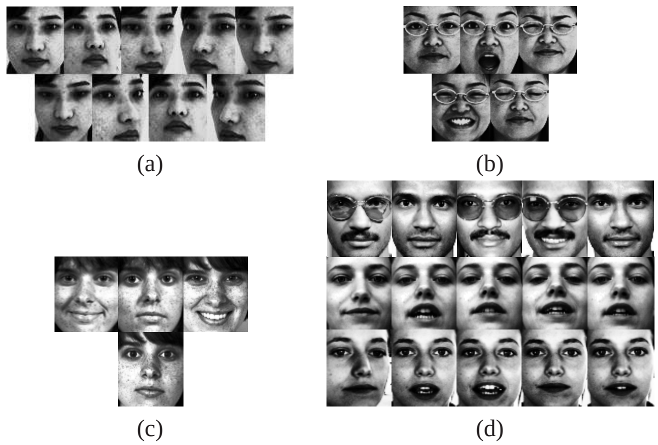

Figure 2. Images used for experiments. (a) First experiment. First row: training set, second row: test set. (b) Second experiment. First row: training set, second row: test set. (c) Third experiment. First row: training set, the test set #1 contains the middle image of the first row (photo taken on 10/31/1994) and the test set #2 contains the second row image (photo taken on 05/21/1996). (d) Fourth experiment. First row: training set from the FERET database. Second (resp. third) row: an extract of a sequence from the test set #1 (resp. #2) from the BioId database.

and test sets present similar lighting conditions, and neutral facial expressions. The test set is compared to the training set. Figure 3.a shows that, for any number k of projection vectors varying from 1 to 15, 2DO-LDA provides better recognition rates than 2D-PCA. The projection vectors of both methods have the same length (here $w = 65$). The best recognition rate for 2DO-LDA (94,4%) is obtained with $k = 8$ projection vectors, and is 2,1% superior to the best recognition rate for 2D-PCA, obtained with $k = 9$. The PCA+LDA algorithm is computed from a sufficient number of 200 principal components. The best results for LDA are obtained with $k = 40$ projection vectors of length $75 \cdot 65 = 4875$ pixels. Figure 3.b shows that 2DO-LDA gives better results than both 2D-PCA and LDA methods, at the first rank as well as at higher ranks. Therefore, 2DO-LDA appears to be the most robust to head pose changes.

The second experiment (see Figure 2.b) is performed on a subspace of the Asian Face Image Database PF01, with similar lighting conditions and frontal head poses. The training set contains 321 images, i.e. three views per person. One corresponds to a neutral expression; the two others are chosen randomly among the four expressions available in the database: happy, surprised, irritated and closed eyes. The test set contains the two remaining facial expressions, for each person. The test set is compared to the training set. From Figure 3.c we can see that 2DO-LDA gives better results than 2D-PCA with less projection vectors (up to 19,7% of difference with $k = 6$). Figure 3.d shows that

2DO-LDA gives much better recognition rates than both 2D-PCA and LDA methods for a rank varying from 1 to 10 (the mean improvement on the 7 first ranks compared to LDA is about 12%). We have observed that about 68% of the misclassifications of 2D-PCA correspond to matching different persons with the same facial expression, while this kind of errors is involved in only 52% of the fewer misclassifications of 2DO-LDA. Therefore, the efficiency of 2DO-LDA is explained by a better ability to deal with facial expression changes. Indeed, these variations constitute most of the within-class scatter of the training set, which is minimized by 2DO-LDA when applying the criterion 4, while 2D-PCA maximizes the total scatter (containing the within-class scatter).

The third experiment (see Figure 2.c) is performed on the FERET database. The training set contains 666 images of 152 persons. The number of images per person is variable, but always larger than three. For each person, the multiple views are taken on different days, under different lighting conditions. The time interval between two views of the same person can be long (from a few days to almost three years), thus one person can wear eyeglasses or a beard on a photo and not on another one (see first row of Figure 2.d). There are two test sets. The first one contains one image per person, taken from the training set. Test set #2 also contains one image per person, taken another day and not belonging to the training set. Lighting conditions are very different and time delay may be long from one test set to another. All the images used for this experiment contain near-frontal head pose; the facial expression can either be neutral or smiling. Test set #2 is then compared to test set #1. Figure 3.e shows that for both 2DO-LDA and 2D-PCA the recognition rates are low (inferior to 50%), which can be explained by the important dissimilarities between the two test sets. However, 2DO-LDA achieves better recognition rates than 2D-PCA (the best recognition rate for 2DO-LDA is achieved with only 5 2DO-DCs against 13 projection vectors for 2D-PCA, and is 5,3% better). The PCA+LDA algorithm is computed from 500 principal components. From Figure 3.f we can conclude that 2DO-LDA provides better results than 2D-PCA and LDA at the first rank as well as at higher ranks. The storage gain compared to LDA is important given that the best recognition rates for LDA are achieved with 100 projection vectors of length 4875 pixels against only 5 projection vectors of length 65 pixels for 2DO-LDA. Moreover, 2DO-LDA reaches 5,5% mean recognition rate improvement over LDA at the first five ranks.

The last experiment (see Figure 2.d) is performed on both the FERET and BioId databases. The training set and test set #1 of the previous experiment constitute the training set, that contains consequently 818 images from 152 persons. Two test sets are taken from the BioId database. Each one contains 173 images of 18 persons, taken from different video sequences. There are important variations in lighting conditions, facial expression and head pose, within and between the test sets. First, test set #2 is compared on an image-to-image basis to test set #1, and a recognition rate is obtained, as for previous experiments. Given that the two sets contain very different illumination conditions and head poses, the recognition rates are inferior to 65% for both 2DO-LDA and 2D-PCA techniques (see Figure 3.g). However, 2DO-LDA achieves much better recognition rates than 2D-PCA, with 15% difference between their respective maxima. The PCA+LDA algorithm is performed from 600 principal components. From Figure 3.h we can conclude that 2DO-LDA provides better recognition rates than 2D-PCA and LDA at the first rank as well as at higher ranks. The storage gain compared to LDA is very important: for LDA the best recognition rate is achieved with 550 projection vectors of length 4875 pixels against only 5 2DO-DCs of length 65 pixels for 2DO-LDA. Moreover, 2DO-LDA reaches 17.6% mean recognition rate improvement over LDA at the first five ranks. The test sets are then compared on a sequence-to-sequence basis, applying a majority voting scheme to the comparion results obtained previously. 2DO-LDA recognizes 11 sequences from 18, while 2D-PCA recognizes at most 8 sequences.

4. Conclusion

In this paper, we have proposed a new class-based projection method, called 2DO-LDA, that can be successfully applied to face recognition. This technique, by maximizing a generalized Fisher criterion computed directly from matrices of face images, constructs a discriminant projection matrix.

2DO-LDA is numerically more stable and allows an important storage gain in comparison with the usual LDA method. It has already been shown [10, 8] that 2D-PCA outperforms the traditional PCA method. In this paper, we have shown on various databases that 2DO-LDA is more efficient and robust to variations in lighting conditions, facial expression and head pose than both 2D-PCA and LDA.

Acknowledgments

Portions of the research in this paper use the FERET database of facial images collected under the FERET program.

References

[1] P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman, *Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection*, Special Theme Issue on Face and Gesture Recognition, IEEE Trans. on Pattern Analysis and Machine Intelligence, 19(7), 711–720, July 1997.

[2] B.W. Hwang, M.C. Roh and S.W. Lee, *Performance Evaluation of Face Recognition Algorithms on Asian Face Database*, 6^{th} IEEE Int. Conf. on Automatic Face and Gesture

Recognition, 278–283, Seoul, Korea, May 2004.

[3] O. Jesorsky, K.J. Kirchberg and R.W. Frischholz, *Robust Face Detection Using the Hausdorff Distance*, In Proc. 3^{rd} International Conference on Audio- and Video-based Biometric Person Authentication, Springer, Lecture Notes in Computer Science, LNCS-2091, 90–95, Halmstad, Sweden, June 2001.

[4] P.J. Phillips, H. Wechsler, J. Huang, P. Rauss, *The FERET Database and Evaluation Procedure for Face Recognition Algorithms*, Image and Vision Computing, 16(5), 295–306, 1998.

[5] D.J. Swets and J. Weng, *Using Discriminant Eigenfeatures for Image Retrieval*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(8), 831–836, 1996.

[6] L. Sirovitch and M. Kirby, *A Low-Dimensional Procedure for the Characterization of Human Faces*, Journ. of Optical Society of America A, 4(3), 519–524, 1987.

[7] M. Turk and A. Pentland, *Face recognition using eigenfaces*, Proc. IEEE Conference on Computer Vision and Pattern Recognition, 586–591, Maui, Hawaii, 1991.

[8] M. Visani, C. Garcia and C. Laurent, *Comparing Robustness of Two-Dimensional PCA and Eigenfaces for Face Recognition*, *To appear in* 1^{st} International Conference on Image Analysis and Recognition (ICIAR), September 2004.

[9] W.S. Yambor, B. Draper and R. Beveridge, *Analyzing PCA-based Face Recognition Algorithms: Eigenvector Selection and Distance Measures*, in Empirical Evaluation Methods in Computer Vision, H. Christensen and J. Phillips (eds.), World Scientific Press, Singapore, 2002.

[10] J. Yang, D. Zhang and A.F. Frangi, *Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(1), 131–137, January 2004.

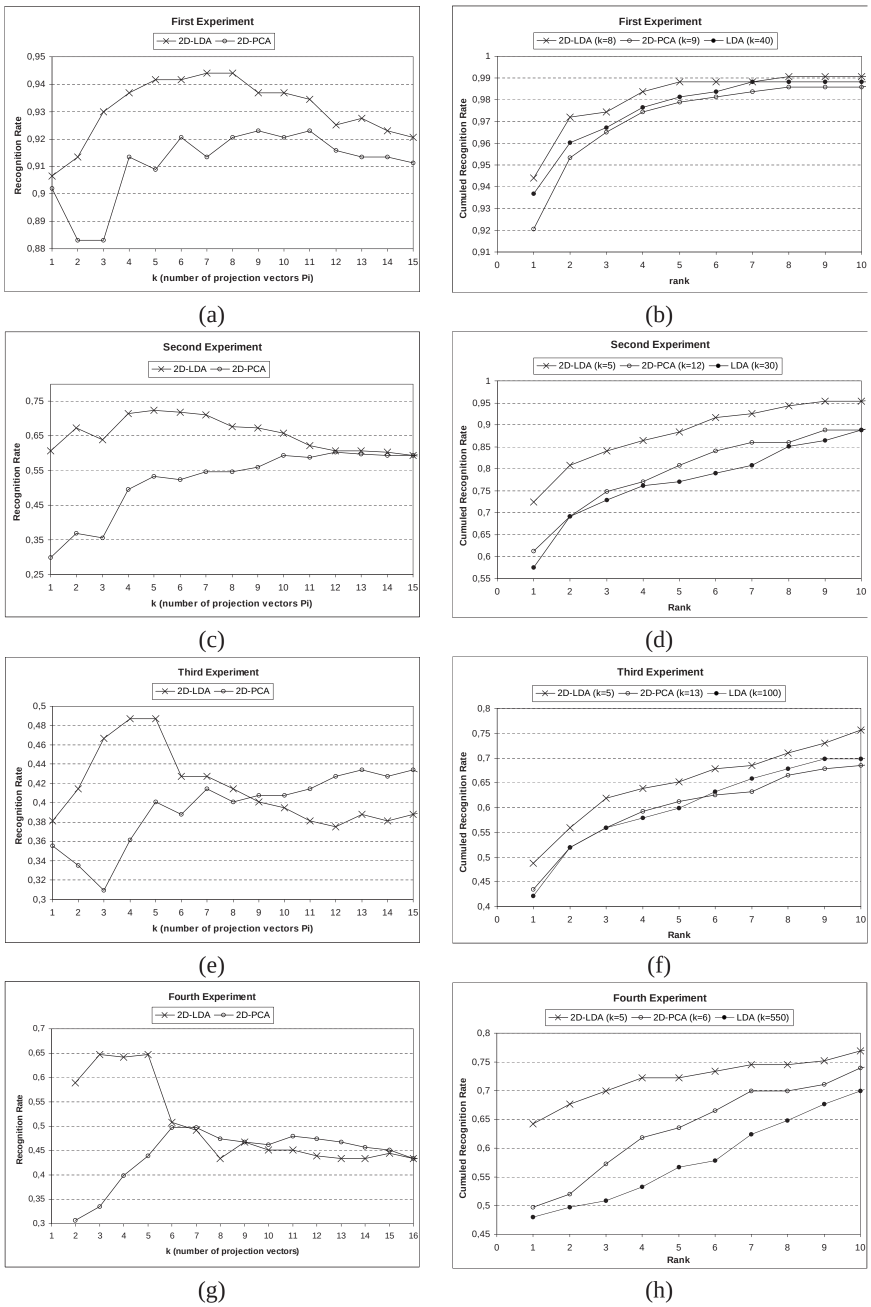

Figure 3. Compared recognition rates of 2DO-LDA and 2D-PCA when varying the number k of projection vectors (first column) and compared CMC curves of 2DO-LDA, 2D-PCA and LDA, for each of the four experiments (second column).